

Closed-Form Factorization of Latent Semantics in GANs

Year: 2020 | Citations: 626 | Authors: Yujun Shen, Bolei Zhou

Abstract

A rich set of interpretable dimensions has been shown to emerge in the latent space of the Generative Adversarial Networks (GANs) trained for synthesizing images. In order to identify such latent dimensions for image editing, previous methods typically annotate a collection of synthesized samples and train linear classifiers in the latent space. However, they require a clear definition of the target attribute as well as the corresponding manual annotations, limiting their applications in practice. In this work, we examine the internal representation learned by GANs to reveal the underlying variation factors in an unsupervised manner. In particular, we take a closer look into the generation mechanism of GANs and further propose a closedform factorization algorithm for latent semantic discovery by directly decomposing the pre-trained weights. With a lightning-fast implementation, our approach is capable of not only finding semantically meaningful dimensions comparably to the state-of-the-art supervised methods, but also resulting in far more versatile concepts across multiple GAN models trained on a wide range of datasets.¹